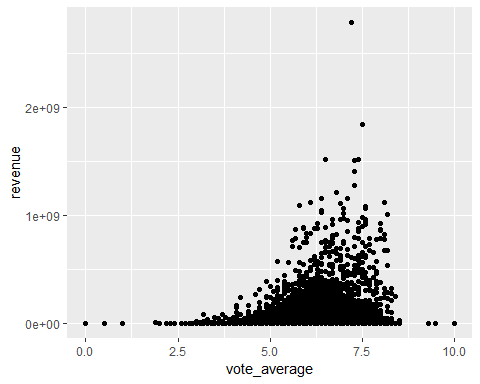
library(ggplot2)  
library(readxl)  
movies.dataset <- read\_excel("tmdb\_5000\_movies.xls")  
summary(movies.dataset)

## budget genres homepage   
## Min. : 0 Length:4803 Length:4803   
## 1st Qu.: 790000 Class :character Class :character   
## Median : 15000000 Mode :character Mode :character   
## Mean : 29045040   
## 3rd Qu.: 40000000   
## Max. :380000000   
##   
## id keywords original\_language original\_title   
## Min. : 5 Length:4803 Length:4803 Length:4803   
## 1st Qu.: 9014 Class :character Class :character Class :character   
## Median : 14629 Mode :character Mode :character Mode :character   
## Mean : 57166   
## 3rd Qu.: 58611   
## Max. :459488   
##   
## overview popularity production\_companies  
## Length:4803 Min. : 0.000 Length:4803   
## Class :character 1st Qu.: 4.668 Class :character   
## Mode :character Median : 12.922 Mode :character   
## Mean : 21.492   
## 3rd Qu.: 28.314   
## Max. :875.581   
##   
## production\_countries release\_date revenue   
## Length:4803 Min. :1916-09-04 00:00:00 Min. :0.000e+00   
## Class :character 1st Qu.:1999-07-14 00:00:00 1st Qu.:0.000e+00   
## Mode :character Median :2005-10-03 00:00:00 Median :1.917e+07   
## Mean :2002-12-27 23:45:54 Mean :8.226e+07   
## 3rd Qu.:2011-02-16 00:00:00 3rd Qu.:9.292e+07   
## Max. :2017-02-03 00:00:00 Max. :2.788e+09   
## NA's :1   
## runtime spoken\_languages status tagline   
## Min. : 0.0 Length:4803 Length:4803 Length:4803   
## 1st Qu.: 94.0 Class :character Class :character Class :character   
## Median :103.0 Mode :character Mode :character Mode :character   
## Mean :106.9   
## 3rd Qu.:118.0   
## Max. :338.0   
## NA's :2   
## title vote\_average vote\_count   
## Length:4803 Min. : 0.000 Min. : 0.0   
## Class :character 1st Qu.: 5.600 1st Qu.: 54.0   
## Mode :character Median : 6.200 Median : 235.0   
## Mean : 6.092 Mean : 690.2   
## 3rd Qu.: 6.800 3rd Qu.: 737.0   
## Max. :10.000 Max. :13752.0   
##

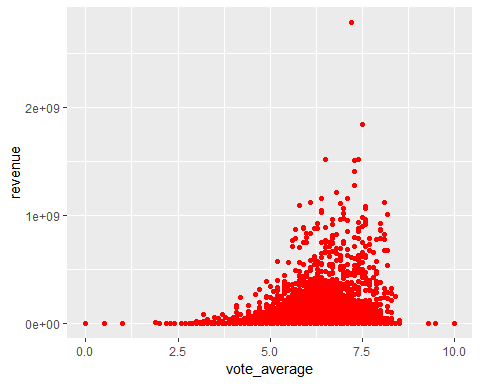
str(movies.dataset)

## Classes 'tbl\_df', 'tbl' and 'data.frame': 4803 obs. of 20 variables:  
## $ budget : num 2.37e+08 3.00e+08 2.45e+08 2.50e+08 2.60e+08 2.58e+08 2.60e+08 2.80e+08 2.50e+08 2.50e+08 ...  
## $ genres : chr "[{\"id\": 28, \"name\": \"Action\"}, {\"id\": 12, \"name\": \"Adventure\"}, {\"id\": 14, \"name\": \"Fantasy\"}"| \_\_truncated\_\_ "[{\"id\": 12, \"name\": \"Adventure\"}, {\"id\": 14, \"name\": \"Fantasy\"}, {\"id\": 28, \"name\": \"Action\"}]" "[{\"id\": 28, \"name\": \"Action\"}, {\"id\": 12, \"name\": \"Adventure\"}, {\"id\": 80, \"name\": \"Crime\"}]" "[{\"id\": 28, \"name\": \"Action\"}, {\"id\": 80, \"name\": \"Crime\"}, {\"id\": 18, \"name\": \"Drama\"}, {\"i"| \_\_truncated\_\_ ...  
## $ homepage : chr "http://www.avatarmovie.com/" "http://disney.go.com/disneypictures/pirates/" "http://www.sonypictures.com/movies/spectre/" "http://www.thedarkknightrises.com/" ...  
## $ id : num 19995 285 206647 49026 49529 ...  
## $ keywords : chr "[{\"id\": 1463, \"name\": \"culture clash\"}, {\"id\": 2964, \"name\": \"future\"}, {\"id\": 3386, \"name\": \""| \_\_truncated\_\_ "[{\"id\": 270, \"name\": \"ocean\"}, {\"id\": 726, \"name\": \"drug abuse\"}, {\"id\": 911, \"name\": \"exotic "| \_\_truncated\_\_ "[{\"id\": 470, \"name\": \"spy\"}, {\"id\": 818, \"name\": \"based on novel\"}, {\"id\": 4289, \"name\": \"secr"| \_\_truncated\_\_ "[{\"id\": 849, \"name\": \"dc comics\"}, {\"id\": 853, \"name\": \"crime fighter\"}, {\"id\": 949, \"name\": \""| \_\_truncated\_\_ ...  
## $ original\_language : chr "en" "en" "en" "en" ...  
## $ original\_title : chr "Avatar" "Pirates of the Caribbean: At World's End" "Spectre" "The Dark Knight Rises" ...  
## $ overview : chr "In the 22nd century, a paraplegic Marine is dispatched to the moon Pandora on a unique mission, but becomes tor"| \_\_truncated\_\_ "Captain Barbossa, long believed to be dead, has come back to life and is headed to the edge of the Earth with W"| \_\_truncated\_\_ "A cryptic message from Bondâ\200\231s past sends him on a trail to uncover a sinister organization. While M bat"| \_\_truncated\_\_ "Following the death of District Attorney Harvey Dent, Batman assumes responsibility for Dent's crimes to protec"| \_\_truncated\_\_ ...  
## $ popularity : num 150.4 139.1 107.4 112.3 43.9 ...  
## $ production\_companies: chr "[{\"name\": \"Ingenious Film Partners\", \"id\": 289}, {\"name\": \"Twentieth Century Fox Film Corporation\", \"| \_\_truncated\_\_ "[{\"name\": \"Walt Disney Pictures\", \"id\": 2}, {\"name\": \"Jerry Bruckheimer Films\", \"id\": 130}, {\"name"| \_\_truncated\_\_ "[{\"name\": \"Columbia Pictures\", \"id\": 5}, {\"name\": \"Danjaq\", \"id\": 10761}, {\"name\": \"B24\", \"id\": 69434}]" "[{\"name\": \"Legendary Pictures\", \"id\": 923}, {\"name\": \"Warner Bros.\", \"id\": 6194}, {\"name\": \"DC E"| \_\_truncated\_\_ ...  
## $ production\_countries: chr "[{\"iso\_3166\_1\": \"US\", \"name\": \"United States of America\"}, {\"iso\_3166\_1\": \"GB\", \"name\": \"United Kingdom\"}]" "[{\"iso\_3166\_1\": \"US\", \"name\": \"United States of America\"}]" "[{\"iso\_3166\_1\": \"GB\", \"name\": \"United Kingdom\"}, {\"iso\_3166\_1\": \"US\", \"name\": \"United States of America\"}]" "[{\"iso\_3166\_1\": \"US\", \"name\": \"United States of America\"}]" ...  
## $ release\_date : POSIXct, format: "2009-12-10" "2007-05-19" ...  
## $ revenue : num 2.79e+09 9.61e+08 8.81e+08 1.08e+09 2.84e+08 ...  
## $ runtime : num 162 169 148 165 132 139 100 141 153 151 ...  
## $ spoken\_languages : chr "[{\"iso\_639\_1\": \"en\", \"name\": \"English\"}, {\"iso\_639\_1\": \"es\", \"name\": \"Espa\\u00f1ol\"}]" "[{\"iso\_639\_1\": \"en\", \"name\": \"English\"}]" "[{\"iso\_639\_1\": \"fr\", \"name\": \"Fran\\u00e7ais\"}, {\"iso\_639\_1\": \"en\", \"name\": \"English\"}, {\"iso\_"| \_\_truncated\_\_ "[{\"iso\_639\_1\": \"en\", \"name\": \"English\"}]" ...  
## $ status : chr "Released" "Released" "Released" "Released" ...  
## $ tagline : chr "Enter the World of Pandora." "At the end of the world, the adventure begins." "A Plan No One Escapes" "The Legend Ends" ...  
## $ title : chr "Avatar" "Pirates of the Caribbean: At World's End" "Spectre" "The Dark Knight Rises" ...  
## $ vote\_average : num 7.2 6.9 6.3 7.6 6.1 5.9 7.4 7.3 7.4 5.7 ...  
## $ vote\_count : num 11800 4500 4466 9106 2124 ...

ggplot(movies.dataset, aes(vote\_average, revenue))+geom\_point()



ggplot(movies.dataset, aes(vote\_average, revenue))+geom\_point(colour="red")



ggplot(movies.dataset, aes(budget, revenue))+geom\_point()+geom\_smooth()

## `geom\_smooth()` using method = 'gam' and formula 'y ~ s(x, bs = "cs")'

